**Project Overview**

The goal of this project was to create a **data warehouse** to analyze coffee shop sales. The data warehouse helps in organizing and storing sales data in a structured way, making it easy to analyze.

**Key Steps in the Project:**

* **Data Sourcing:** We collected raw data from CSV files.
* **ETL (Extract, Transform, Load):** The data was cleaned, formatted, and loaded into tables.
* **Dimensional Modeling:** We designed a database using a **star schema**, making queries faster.
* **Analytical Queries:** We wrote SQL queries to get useful business insights.

We used **Snowflake** as the database platform and **SQL** for working with the data.

**Challenges Faced and How We Solved Them**

|  |  |
| --- | --- |
| **Problem** | **Solution** |
| **Date formats were inconsistent** | We used TO\_DATE() to convert all dates into a standard format. |
| **Some values had extra characters (e.g., "$18.00")** | We used REPLACE() to remove symbols and converted them into numbers. |
| **Missing store details in staff data** | We assigned fixed values for headquarters (999) and warehouse (888). |
| **Some sales transactions had missing product or customer details** | We ran checks to find missing data and ensured only valid records were used. |
| **Duplicate records in the data** | We used DISTINCT to remove duplicates before loading the data. |
| **Queries were slow due to large data volume** | We kept only necessary columns and avoided unnecessary calculations in queries. |

By handling these issues, we ensured the data was **clean, complete, and reliable** for analysis.

**How Well the Dimensional Model Worked**

We used a **star schema** to organize the data. This means:

* The **fact table** (fact\_sales) stores all sales transactions.
* Several **dimension tables** (dim\_customer, dim\_product, dim\_store, dim\_staff, dim\_date) store details related to customers, products, stores, employees, and dates.

**Why This Model Was Effective:**

**Faster Queries:** Since the tables were well-structured, it was easier to get sales insights quickly.  
**Easy to Understand:** Each table had a clear purpose, making it simple to analyze.  
**Better Organization:** Sales data was stored separately from product and customer details, avoiding confusion.

**Scope for Improvement**

**Make the Data More Useful:**

* Add more details about customers, such as their buying preferences.
* Track product prices over time to see how they change.

**Improve the Analysis:**

* Use charts and graphs for better data visualization.
* Compare sales over longer periods instead of just one month.

**Keep the Data Updated:**

* Automate the process of adding new sales records so the data warehouse stays current.
* Regularly check for errors or missing values before using the data.

**Conclusion**

We were able to build data warehouse that organizes coffee shop sales data in a way that makes it easy to analyze. The star schema model worked well, allowing for quick queries and clear relationships between sales and other business details.

With improvements like adding historical tracking, more data sources, and better automation, this system could become even more useful for analyzing long-term business trends.

Overall, this project helped develop skills in data organization, SQL querying, and business analysis, all of which are important for working with large datasets.